

## SEQUENCE LISTING

### (1) GENERAL INFORMATION:

- (i) APPLICANT: De Robertis, Edward M.  
Bouwmeester, Tewis
- (ii) TITLE OF INVENTION: Endoderm, Cardiac and Neural Inducing  
Factors
- (iii) NUMBER OF SEQUENCES: 10
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: Majestic, Parsons, Siebert & Hsue
  - (B) STREET: Four Embarcadero Center, Suite 1100
  - (C) CITY: San Francisco
  - (D) STATE: California
  - (E) COUNTRY: U.S.A.
  - (F) ZIP: 94111-4106
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: Floppy disk
  - (B) COMPUTER: IBM PC compatible
  - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
  - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 08/878,474
  - (B) FILING DATE: 18-JUN-1997
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: US 60/020,150
  - (B) FILING DATE: 20-JUN-1996
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: Siebert, J. Suzanne
  - (B) REGISTRATION NUMBER: 28,758
  - (C) REFERENCE/DOCKET NUMBER: 3100.002US1
- (ix) TELECOMMUNICATION INFORMATION:
  - (A) TELEPHONE: 415/248-5500
  - (B) TELEFAX: 415/362-5418

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 270 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Met Leu Leu Asn Val Leu Arg Ile Cys Ile Ile Val Cys Leu Val Asn  
1 5 10 15  
Asp Gly Ala Gly Lys His Ser Glu Gly Arg Glu Arg Thr Lys Thr Tyr  
20 25 30  
Ser Leu Asn Ser Arg Gly Tyr Phe Arg Lys Glu Arg Gly Ala Arg Arg  
35 40 45  
Ser Lys Ile Leu Leu Val Asn Thr Lys Gly Leu Asp Glu Pro His Ile  
50 55 60  
Gly His Gly Asp Phe Gly Leu Val Ala Glu Leu Phe Asp Ser Thr Arg  
65 70 75 80  
Thr His Thr Asn Arg Lys Glu Pro Asp Met Asn Lys Val Lys Leu Phe  
85 90 95  
Ser Thr Val Ala His Gly Asn Lys Ser Ala Arg Arg Lys Ala Tyr Asn  
100 105 110  
Gly Ser Arg Arg Asn Ile Phe Ser Arg Arg Ser Phe Asp Lys Arg Asn  
115 120 125  
Thr Glu Val Thr Glu Lys Pro Gly Ala Lys Met Phe Trp Asn Asn Phe  
130 135 140  
Leu Val Lys Met Asn Gly Ala Pro Gln Asn Thr Ser His Gly Ser Lys  
145 150 155 160  
Ala Gln Glu Ile Met Lys Glu Ala Cys Lys Thr Leu Pro Phe Thr Gln  
165 170 175  
Asn Ile Val His Glu Asn Cys Asp Arg Met Val Ile Gln Asn Asn Leu  
180 185 190  
Cys Phe Gly Lys Cys Ile Ser Leu His Val Pro Asn Gln Gln Asp Arg  
195 200 205  
Arg Asn Thr Cys Ser His Cys Leu Pro Ser Lys Phe Thr Leu Asn His  
210 215 220

Leu	Thr	Leu	Asn	Cys	Thr	Gly	Ser	Lys	Asn	Val	Val	Lys	Val	Val	Met
225					230					235					240
Met	Val	Glu	Glu	Cys	Thr	Cys	Glu	Ala	His	Lys	Ser	Asn	Phe	His	Gln
				245					250					255	
Thr	Ala	Gln	Phe	Asn	Met	Asp	Thr	Ser	Thr	Thr	Leu	His	His		
			260					265					270		

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1411 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GAATTCCTAA AAGCGGCACA GTGCAGGAAC AGCAAGTCGC TCAGAAACAC TGCAGGGTCT	60
AGATATCATA CAATGTTACT AAATGTACTC AGGATCTGTA TTATCGTCTG CCTTGTGAAT	120
GATGGAGCAG GAAAACACTC AGAAGGACGA GAAAGGACAA AAACATATTC ACTTAACAGC	180
AGAGGTTACT TCAGAAAAGA AAGAGGAGCA CGTAGGAGCA AGATTCTGCT GGTGAATACT	240
AAAGGTCTTG ATGAACCCCA CATTGGGCAT GGTGATTTTG GCTTAGTAGC TGAACTATTT	300
GATTCCACCA GAACACATAC AAACAGAAAA GAGCCAGACA TGAACAAAGT CAAGCTTTTC	360
TCAACAGTTG CCCATGGAAA CAAAAGTGCA AGAAGAAAAG CTTACAATGG TTCTAGAAGG	420
AATATTTTTT CTCGCCGTTT TTTTGATAAA AGAAATACAG AGGTTACTGA AAAGCCTGGT	480
GCCAAGATGT TCTGGAACAA TTTTGTGGTT AAAATGAATG GAGCCCCACA GAATACAAGC	540
CATGGCAGTA AAGCACAGGA AATAATGAAA GAAGCTTGCA AAACCTTGCC CTTCACTCAG	600
AATATTGTAC ATGAAACTG TGACAGGATG GTGATACAGA ACAATCTGTG CTTTGGTAAA	660
TGCATCTCTC TCCATGTTCC AAATCAGCAA GATCGACGAA ATACTTGTTT CCATTGCTTG	720
CCGTCCAAAT TTACCCTGAA CCACCTGACG CTGAATTGTA CTGGATCTAA GAATGTAGTA	780
AAGGTTGTCA TGATGGTAGA GGAATGCACG TGTGAAGCTC ATAAGAGCAA CTTCCACCAA	840

ACTGCACAGT TTAACATGGA TACATCTACT ACCCTGCACC ATTAAAAGGA CTGTCTGCCA 900  
TACAGTATGG AAATGCCCAT TTGTTGGAAT ATTCGTTACA TGCTATGTAT CTAAAGCATT 960  
ATGTTGCCTT CTGTTTCATA TAACCACATG GAATAAGGAT TGTATGAATT ATAATTAACA 1020  
AATGGCATT TGTGTAACAT GCAAGATCTC TGTTCATCA GTTGCAAGAT AAAAGGCAAT 1080  
ATTTGTTTGA CTTTTTCTA CAAAATGAAT ACCCAAATAT ATGATAAGAT AATGGGGTCA 1140  
AAACTGTTAA GGGGTAATGT AATAATAGGG ACTACAACC AATCAGCAGG TATGATTTAC 1200  
TGGTCACCTG TTAAAAGCA AACATCTTAT TGGTTGCTAT GGGTACTGC TTCTGGGCAA 1260  
AATGTGTGCC TCATAGGGGG GTTAGTGTGT TGTGTACTGA ATTAATTGTA TTTATTTTCAT 1320  
TGTTACAATG AAGAGGATGT CTATGTTTAT TTCACTTTTA TTAATGTACA ATAAATGTTC 1380  
TTGTTTCTTT AAAAAAAAAA AAAAATCGA G 1411

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 318 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met	Ser	Arg	Thr	Arg	Lys	Val	Asp	Ser	Leu	Leu	Leu	Leu	Ala	Ile	Pro
1				5					10					15	
Gly	Leu	Ala	Leu	Leu	Leu	Leu	Pro	Asn	Ala	Tyr	Cys	Ala	Ser	Cys	Glu
			20					25					30		
Pro	Val	Arg	Ile	Pro	Met	Cys	Lys	Ser	Met	Pro	Trp	Asn	Met	Thr	Lys
		35					40					45			
Met	Pro	Asn	His	Leu	His	His	Ser	Thr	Gln	Ala	Asn	Ala	Ile	Leu	Ala
	50					55					60				
Ile	Glu	Gln	Phe	Glu	Gly	Leu	Leu	Thr	Thr	Glu	Cys	Ser	Gln	Asp	Leu
65					70					75				80	
Leu	Phe	Phe	Leu	Cys	Ala	Met	Tyr	Ala	Pro	Ile	Cys	Thr	Ile	Asp	Phe
			85						90					95	

Gln His Glu Pro Ile Lys Pro Cys Lys Ser Val Cys Glu Arg Ala Arg  
 100 105 110  
 Ala Gly Cys Glu Pro Ile Leu Ile Lys Tyr Arg His Thr Trp Pro Glu  
 115 120 125  
 Ser Leu Ala Cys Glu Glu Leu Pro Val Tyr Asp Arg Gly Val Cys Ile  
 130 135 140  
 Ser Pro Glu Ala Ile Val Thr Val Glu Gln Gly Thr Asp Ser Met Pro  
 145 150 155 160  
 Asp Phe Ser Met Asp Ser Asn Asn Gly Asn Cys Gly Ser Gly Arg Glu  
 165 170 175  
 His Cys Lys Cys Lys Pro Met Lys Ala Thr Gln Lys Thr Tyr Leu Lys  
 180 185 190  
 Asn Asn Tyr Asn Tyr Val Ile Arg Ala Lys Val Lys Glu Val Lys Val  
 195 200 205  
 Lys Cys His Asp Ala Thr Ala Ile Val Glu Val Lys Glu Ile Leu Lys  
 210 215 220  
 Ser Ser Leu Val Asn Ile Pro Lys Asp Thr Val Thr Leu Tyr Thr Asn  
 225 230 235 240  
 Ser Gly Cys Leu Cys Pro Gln Leu Val Ala Asn Glu Glu Tyr Ile Ile  
 245 250 255  
 Met Gly Tyr Glu Asp Lys Glu Arg Thr Arg Leu Leu Leu Val Glu Gly  
 260 265 270  
 Ser Leu Ala Glu Lys Trp Arg Asp Arg Leu Ala Lys Lys Val Lys Arg  
 275 280 285  
 Trp Asp Gln Lys Leu Arg Arg Pro Arg Lys Ser Lys Asp Pro Val Ala  
 290 295 300  
 Pro Ile Pro Asn Lys Asn Ser Asn Ser Arg Gln Ala Arg Ser  
 305 310 315

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1875 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAATTCCTT TCACACAGGA CTCCTGGCAG AGGTGAATGG TTAGCCCTAT GGATTGTT 60  
TGTTGATTTT GACACATGAT TGATTGCTTT CAGATAGGAT TGAAGGACTT GGATTTTAT 120  
CTAATTCTGC ACTTTTAAAT TATCTGAGTA ATTGTTCAAT TTGTATTGGA TGGGACTAAA 180  
GATAAACTTA ACTCCTTGCT TTTGACTTGC CCATAAACTA TAAGGTGGGG TGAGTTGTAG 240  
TTGCTTTTAC ATGTGCCAG ATTTTCCCTG TATTCCCTGT ATTCCCTCTA AAGTAAGCCT 300  
ACACATACAG GTTGGGCAGA ATAACAATGT CTCGAACAAG GAAAGTGGAC TCATTACTGC 360  
TACTGGCCAT ACCTGGACTG GCGCTTCTCT TATTACCCAA TGCTTACTGT GCTTCGTGTG 420  
AGCCTGTGCG GATCCCCATG TGCAAATCTA TGCCATGGAA CATGACCAAG ATGCCCAACC 480  
ATCTCCACCA CAGCACTCAA GCCAATGCCA TCCTGGCAAT TGAACAGTTT GAAGGTTTGC 540  
TGACCACTGA ATGTAGCCAG GACCTTTTGT TCTTTCTGTG TGCCATGTAT GCCCCATTT 600  
GTACCATCGA TTTCCAGCAT GAACCAATTA AGCCTTGCAA GTCCGTGTGC GAAAGGGCCA 660  
GGGCCGGCTG TGAGCCCATT CTCATAAAGT ACCGGCACAC TTGGCCAGAG AGCCTGGCAT 720  
GTGAAGAGCT GCGCGTATAT GACAGAGGAG TCTGCATCTC CCCAGAGGCT ATCGTCACAG 780  
TGGAACAAGG AACAGATTCA ATGCCAGACT TCTCCATGGA TTCAAACAAT GGAAATTGCG 840  
GAAGCGGCAG GGAGCACTGT AAATGCAAGC CCATGAAGGC AACCCAAAAG ACGTATCTCA 900  
AGAATAATTA CAATTATGTA ATCAGAGCAA AAGTGAAAGA GGTGAAAGTG AAATGCCACG 960  
ACGCAACAGC AATTGTGGAA GTAAAGGAGA TTCTCAAGTC TTCCCTAGTG AACATTCTTA 1020  
AAGACACAGT GACACTGTAC ACCAACTCAG GCTGCTTGTG CCCCAGCTT GTTGCCAATG 1080  
AGGAATACAT AATTATGGGC TATGAAGACA AAGAGCGTAC CAGGCTTCTA CTAGTGAAG 1140  
GATCCTTGGC CGAAAAATGG AGAGATCGTC TTGCTAAGAA AGTCAAGCGC TGGGATCAAA 1200  
AGCTTCGACG TCCCAGGAAA AGCAAAGACC CCGTGGCTCC AATTCCCAAC AAAACAGCA 1260  
ATTCCAGACA AGCGCGTAGT TAGACTAACG GAAAGGTGTA TGGAAACTCT ATGGACTTTG 1320  
AAACTAAGAT TTGCATTGTT GGAAGAGCAA AAAAGAAATT GCACTACAGC ACGTTATATT 1380  
CTATTGTTTA CTACAAGAAG CTGGTTTAGT TGATTGTAGT TCTCCTTTCC TTCTTTTTTT 1440

TTATAACTAT ATTTGCACGT GTTCCCAGGC AATTGTTTTA TTCAACTTCC AGTGACAGAG	1500
CAGTGACTGA ATGTCTCAGC CTAAAGAAGC TCAATTCATT TCTGATCAAC TAATGGTGAC	1560
AAGTGTTTGA TACTTGGGGA AAGTGA ACTA ATTGCAATGG TAAATCAGAG AAAAGTTGAC	1620
CAATGTTGCT TTTCTGTAG ATGAACAAGT GAGAGATCAC ATTTAAATGA TGATCACTTT	1680
CCATTTAATA CTTTCAGCAG TTTTAGTTAG ATGACATGTA GGATGCACCT AAATCTAAAT	1740
ATTTTATCAT AAATGAAGAG CTGGTTTAGA CTGTATGGTC ACTGTTGGGA AGGTAAATGC	1800
CTACTTTGTC AATTCTGTTT TAAAATTGC CTAAATAAAT ATTAAGTCCT AAATAAAAAA	1860
AAAAAAAAAA AAAAA	1875

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 979 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Leu	Leu	Leu	Phe	Arg	Ala	Ile	Pro	Met	Leu	Leu	Leu	Gly	Leu	Met	1	5	10	15
Val	Leu	Gln	Thr	Asp	Cys	Glu	Ile	Ala	Gln	Tyr	Tyr	Ile	Asp	Glu	Glu	20	25	30	
Glu	Pro	Pro	Gly	Thr	Val	Ile	Ala	Val	Leu	Ser	Gln	His	Ser	Ile	Phe	35	40	45	
Asn	Thr	Thr	Asp	Ile	Pro	Ala	Thr	Asn	Phe	Arg	Leu	Met	Lys	Gln	Phe	50	55	60	
Asn	Asn	Ser	Leu	Ile	Gly	Val	Arg	Glu	Ser	Asp	Gly	Gln	Leu	Ser	Ile	65	70	75	80
Met	Glu	Arg	Ile	Asp	Arg	Glu	Gln	Ile	Cys	Arg	Gln	Ser	Leu	His	Cys	85	90	95	
Asn	Leu	Ala	Leu	Asp	Val	Val	Ser	Phe	Ser	Lys	Gly	His	Phe	Lys	Leu	100	105	110	

**THE** **NEW** **YORK** **PUBLIC** **LIBRARY**







Val Ala Asn Arg Met His Ala Glu Tyr Glu Arg Asp Leu Val Asn Arg  
 930 935 940

Ser Ala Thr Leu Ser Pro Gln Arg Ser Ser Ser Arg Tyr Gln Glu Phe  
 945 950 955 960

Asn Tyr Ser Pro Gln Ile Ser Arg Gln Leu His Pro Ser Glu Ile Ala  
 965 970 975

Thr Thr Phe

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3655 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTC	CCAG	AGATGA	ACTC	CTTGAG	ATTG	TTTTAA	ATGA	CTGCAG	GTCT	GGAAGG	ATTC	60
ACATTG	CCAC	ACTGTT	TCTA	GGCATG	AAAA	AACTGC	AAAGT	TTCAAC	TTTG	TTTTTG	GGTG	120
AACTTT	GATT	CTTCAAG	ATG	CTGCTT	CTCT	TCAGAG	CCAT	TCCAAT	GCTG	CTGTTG	GGGAC	180
TGATGG	TTTT	ACAAAC	AGAC	TGTGAA	ATTG	CCCAGT	ACTA	CATAGAT	GAA	GAAGA	ACCCC	240
CTGGCA	CTGT	AATTGC	AGTG	TTGTCAC	AAC	ACTCCAT	ATT	TAACACT	ACA	GATATAC	CTG	300
CAACCA	ATTT	CCGTCT	AATG	AAGCAAT	TTA	ATAATT	CCCT	TATCGG	AGTC	CGTGAG	AGTG	360
ATGGGC	CAGCT	GAGCAT	CATG	GAGAGG	ATTG	ACCGGG	AGCA	AATCTG	CAGG	CAGTCC	CCTC	420
ACTGCA	ACCT	GGCTTT	GGAT	GTGGTC	AGCT	TTTCCA	AAGG	AACTTCA	AG	CTTCTG	AACG	480
TGAAAG	TGGA	GGTGAG	AGAC	ATTAAT	GACC	ATAGCC	CCTCA	CTTTCC	CAGT	GAAATA	AATGC	540
ATGTGG	AGGT	GTCTGA	AAGT	TCCTCT	GTGG	GCACCAG	GAT	TCCTTT	AGAA	ATTGCA	ATAG	600
ATGAAG	ATGT	TGGGTCC	AAC	TCCATCC	AGA	ACTTTC	AGAT	CTCAA	ATAAT	AGCCACT	TCA	660
GCATTG	ATGT	GCTAAC	CAGA	GCAGAT	GGGG	TGAAAT	ATGC	AGATTT	AGTC	TTAATG	AGAG	720
AACTGG	ACAG	GGAAAT	CCAG	CCAACAT	ACA	TAATGG	AGCT	ACTAGC	AATG	GATGGGG	GTG	780

TACCATCACT	ATCTGGTACT	GCAGTGGTTA	ACATCCGAGT	CCTGGACTTT	AATGATAACA	840
GCCCAGTGTT	TGAGAGAAGC	ACCATTGCTG	TGGACCTAGT	AGAGGATGCT	CCTCTGGGAT	900
ACCTTTTGTT	GGAGTTACAT	GCTACTGACG	ATGATGAAGG	AGTGAATGGA	GAAATTGTTT	960
ATGGATTGAG	CACTTTGGCA	TCTCAAGAGG	TACGTCAGCT	ATTTAAAATT	AACTCCAGAA	1020
CTGGCAGTGT	TACTCTTGAA	GGCCAAGTTG	ATTTTGAGAC	CAAGCAGACT	TACGAATTTG	1080
AGGTACAAGC	CCAAGATTTG	GGCCCCAACC	CACTGACTGC	TACTTGTAAG	GTAAGTGTTC	1140
ATATACTTGA	TGTAAATGAT	AATACCCAG	CCATCACTAT	TACCCCTCTG	ACTACTGTAA	1200
ATGCAGGAGT	TGCCTATATT	CCAGAAACAG	CCACAAAGGA	GAAGTTTATA	GCTCTGATCA	1260
GCACTACTGA	CAGAGCCTCT	GGATCTAATG	GACAAGTTCT	CTGTACTCTT	TATGGACATG	1320
AGCACTTTAA	ACTACAGCAA	GCTTATGAGG	ACAGTTACAT	GATAGTTACC	ACCTCTACTT	1380
TAGACAGGGA	AAACATAGCA	GCGTACTCTT	TGACAGTAGT	TGCAGAAGAC	CTTGGCTTCC	1440
CCTCATTGAA	GACCAAAAAG	TACTACACAG	TCAAGGTTAG	TGATGAGAAT	GACAATGCAC	1500
CTGTATTTTC	TAAACCCAG	TATGAAGCTT	CTATTCTGGA	AAATAATGCT	CCAGGCTCTT	1560
ATATAACTAC	AGTGATAGCC	AGAGACTCTG	ATAGTGATCA	AAATGGCAAA	GTAAATTACA	1620
GACTTGTGGA	TGCAAAAGTG	ATGGGCCAGT	CACTAACAAC	ATTTGTTTCT	CTTGATGCGG	1680
ACTCTGGAGT	ATTGAGAGCT	GTTAGGTCTT	TAGACTATGA	AAAGTTTAAA	CAACTGGATT	1740
TTGAAATTGA	AGCTGCAGAC	AATGGGATCC	CTCAACTCTC	CACTCGCGTT	CAACTAAATC	1800
TCAGAATAGT	TGATCAAAAT	GATAATTGCC	CTGTGATAAC	TAATCCTCTT	CTTAATAATG	1860
GCTCGGGTGA	AGTTCTGCTT	CCCATCAGCG	CTCCTCAAAA	CTATTTAGTT	TTCCAGCTCA	1920
AAGCCGAGGA	TTCAGATGAA	GGGCACAACT	CCCAGCTGTT	CTATACCATA	CTGAGAGATC	1980
CAAGCAGATT	GTTTGCCATT	AACAAAGAAA	GTGGTGAAGT	GTTCTTGAAA	AAACAATTAA	2040
ACTCTGACCA	TTCAGAGGAC	TTGAGCATAG	TAGTTGCAGT	GTATGACTTG	GGAAGACCTT	2100
CATTATCCAC	CAATGCTACA	GTTAAATTCA	TCCTCACCAG	CTCTTTTCCT	TCTAACGTTG	2160
AAGTCGTTAT	TTTGCAACCA	TCTGCAGAAG	AGCAGCACCA	GATCGATATG	TCCATTATAT	2220
TCATTGCAGT	GCTGGCTGGT	GGTTGTGCTT	TGCTACTTTT	GGCCATCTTT	TTTGTGGCCT	2280
GTAAGTGTAA	AAAGAAAGCT	GGTGAATTTA	AGCAGGTACC	TGAACAACAT	GGAACATGCA	2340

ATGAAGAACG	CCTGTTAAGC	ACCCCATCTC	CCCAGTCGGT	CTCTTCTTCT	TTGTCTCAGT	2400
CTGAGTCATG	CCAACCTCTC	ATCAATACTG	AATCTGAGAA	TTGCAGCGTG	TCCTCTAACC	2460
AAGAGCAGCA	TCAGCAAACA	GGCATAAAGC	ACTCCATCTC	TGTACCATCT	TATCACACAT	2520
CTGGTTGGCA	CCTGGACAAT	TGTGCAATGA	GCATAAGTGG	ACATTCTCAC	ATGGGGCACA	2580
TTAGTACAAA	GGACAGTGGC	AAAGGAGATA	GTGACTTCAA	TGACAGTGAC	TCTGATACTA	2640
GTGGAGAATC	ACAAAAGAAG	AGCATTGAGC	AGCCAATGCA	GGCACAAGCC	AGTGCTCAAT	2700
ACACAGATGA	ATCAGCAGGG	TTCCGACATG	CCGATAACTA	TTTCAGCCAC	CGAATCAACA	2760
AGGGTCCAGA	AAATGGGAAC	TGCACATTGC	AATATGAAAA	GGGCTATAGA	CTGTCTTACT	2820
CTGTAGCTCC	TGCTCATTAC	AATACCTACC	ATGCAAGAAT	GCCTAACCTG	CACATACCGA	2880
ACCATACCCT	TAGAGACCCT	TATTACCATA	TCAATAATCC	TGTTGCTAAT	CGGATGCACG	2940
CGGAATATGA	AAGAGATTTA	GTCAACAGAA	GTGCAACGTT	ATCTCCGCAG	AGATCGTCTA	3000
GCAGATACCA	AGAATTCAAT	TACAGTCCGC	AGATATCAAG	ACAGCTTCAT	CCTTCAGAAA	3060
TTGCTACAAC	CTTTTAATCA	TTAGGCATGC	AAGTGAGAAT	GCACAAAGGC	AAGTGCTTTA	3120
GCATGAAAGC	TAAATATATG	GAGTCTCCCC	TTTCCCTCTG	ATGGATGGGG	GGAGACACAG	3180
GACAGTGCAT	AAATATACAG	CTGCTTTCTA	TTTGCATTTT	ACTTGGAAT	TTTTTGTTTT	3240
TTTTACATAT	TTATTTTTCC	TGAATTGAAT	GTGACATTGT	CCTGTCACCT	AACTAGCAAT	3300
TAAATCCACA	GACCTACAGT	CAAATATTTG	AGGGCCCCTG	AAACAGCACA	TCAGTCAGGA	3360
CCTAAAGTGG	CCTTTTTTACT	TTTAGCAGCT	CCTGGGTCTG	CCCTCTGTGT	TAATCAGCCC	3420
CTGGTCAAGT	CCTGAGTAGG	ATCATGGCGT	TTTTATATGC	ATCTCACCTA	CTTTGGACGT	3480
GATTTACACA	TAATAGGAAA	CGCTTGTTT	CAGTGAAGTC	TGTGTTGTAT	ATATTCTGTT	3540
ATATACACGC	ATTTTGTGTT	TGTGTATATA	TTTCAAGTCC	ATTCAGATAT	GTGTATATAG	3600
TGCAGACCTT	GTAAATTAAA	TATTCTGATA	CTTTTTCCTC	AATAAATATT	TAAAT	3655

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 323 amino acids
- (B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Met Val Cys Cys Gly Pro Gly Arg Met Leu Leu Gly Trp Ala Gly Leu  
1 5 10 15  
Leu Val Leu Ala Ala Leu Cys Leu Leu Gln Val Pro Gly Ala Gln Ala  
20 25 30  
Ala Ala Cys Glu Pro Val Arg Ile Pro Leu Cys Lys Ser Leu Pro Trp  
35 40 45  
Asn Met Thr Lys Met Pro Asn His Leu His His Ser Thr Gln Ala Asn  
50 55 60  
Ala Ile Leu Ala Met Glu Gln Phe Glu Gly Leu Leu Gly Thr His Cys  
65 70 75 80  
Ser Pro Asp Leu Leu Phe Phe Leu Cys Ala Met Tyr Ala Pro Ile Cys  
85 90 95  
Thr Ile Asp Phe Gln His Glu Pro Ile Lys Pro Cys Lys Ser Val Cys  
100 105 110  
Glu Arg Ala Arg Gln Gly Cys Glu Pro Ile Leu Ile Lys Tyr Arg His  
115 120 125  
Ser Trp Pro Glu Ser Leu Ala Cys Asp Glu Leu Pro Val Tyr Asp Arg  
130 135 140  
Gly Val Cys Ile Ser Pro Glu Ala Ile Val Thr Ala Asp Gly Ala Asp  
145 150 155 160  
Phe Pro Met Asp Ser Ser Thr Gly His Cys Arg Gly Ala Ser Ser Glu  
165 170 175  
Arg Cys Lys Cys Lys Pro Val Arg Ala Thr Gln Lys Thr Tyr Phe Arg  
180 185 190  
Asn Asn Tyr Asn Tyr Val Ile Arg Ala Lys Val Lys Glu Val Lys Met  
195 200 205  
Lys Cys His Asp Val Thr Ala Val Val Glu Val Lys Glu Ile Leu Lys  
210 215 220  
Ala Ser Leu Val Asn Ile Pro Arg Asp Thr Val Asn Leu Tyr Thr Thr  
225 230 235 240

Ser Gly Cys Leu Cys Pro Pro Leu Thr Val Asn Glu Glu Tyr Val Ile  
245 250 255

Met Gly Tyr Glu Asp Glu Glu Arg Ser Arg Leu Leu Leu Val Glu Gly  
260 265 270

Ser Ile Ala Glu Lys Trp Lys Asp Arg Leu Gly Lys Lys Val Lys Arg  
275 280 285

Trp Asp Met Lys Leu Arg His Leu Gly Leu Gly Lys Thr Asp Ala Ser  
290 295 300

Asp Ser Thr Gln Asn Gln Lys Ser Gly Arg Asn Ser Asn Pro Arg Pro  
305 310 315 320

Ala Arg Ser

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2176 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AAGCCTGGGA CCATGGTCTG CTGCGGCCCG GGACGGATGC TGCTAGGATG GGCCGGGTTC	60
CTAGTCCTGG CTGCTCTCTG CCTGCTCCAG GTGCCCCGAG CTCAGGCTGC AGCCTGTGAG	120
CCTGTCCGCA TCCCGCTGTG CAAGTCCCTT CCCTGGAACA TGACCAAGAT GCCCAACCAC	180
CTGCACCACA GCACCCAGGC TAACGCCATC CTGGCCATGG AACAGTTCGA AGGGCTGCTG	240
GGCACCCACT GCAGCCCGGA TCTTCTCTTC TTCCTCTGTG CAATGTACGC ACCCATTTGC	300
ACCATCGACT TCCAGCACGA GCCCATCAAG CCCTGCAAGT CTGTGTGTGA GCGCGCCCGA	360
CAGGGCTGCG AGCCCATTCT CATCAAGTAC CGCCACTCGT GGCCGGAAAG CTTGGCCTGC	420
GACGAGCTGC CGGTGTACGA CCGCGGCGTG TGCATCTCTC CTGAGGCCAT CGTCACCGCG	480
GACGGAGCGG ATTTTCCTAT GGATTCAAGT ACTGGACACT GCAGAGGGGC AAGCAGCGAA	540
CGTTGCAAAT GTAAGCCTGT CAGAGCTACA CAGAAGACCT ATTTCCGGAA CAATTACAAC	600

TATGTCATCC	GGGCTAAAGT	TAAAGAGGTA	AAGATGAAAT	GTCATGATGT	GACCGCCGTT	660
GTGGAAGTGA	AGGAAATTCT	AAAGGCATCA	CTGGTAAACA	TTCCAAGGGA	CACCGTCAAT	720
CTTTATACCA	CCTCTGGCTG	CCTCTGTCCT	CCACTTACTG	TCAATGAGGA	ATATGTCATC	780
ATGGGCTATG	AAGACGAGGA	ACGTTCCAGG	TTACTCTTGG	TAGAAGGCTC	TATAGCTGAG	840
AAGTGGAAGG	ATCGGCTTGG	TAAGAAAGTC	AAGCGCTGGG	ATATGAAACT	CCGACACCTT	900
GGACTGGGTA	AAACTGATGC	TAGCGATTCC	ACTCAGAATC	AGAAGTCTGG	CAGGAACTCT	960
AATCCCCGGC	CAGCACGCAG	CTAAATCCTG	AAATGTAAAA	GGCCACACCC	ACGGACTCCC	1020
TTCTAAGACT	GGCGCTGGTG	GACTAACAAA	GGAAAACCGC	ACAGTTGTGC	TCGTGACCGA	1080
TTGTTTACCG	CAGACACCGC	GTGGCTACCG	AAGTTACTTC	CGGTCCCCTT	TCTCCTGCTT	1140
CTTAATGGCG	TGGGGTTAGA	TCCTTTAATA	TGTTATATAT	TCTGTTTCAT	CAATCACGTG	1200
GGGACTGTTC	TTTTGCAACC	AGAATAGTAA	ATTAAATATG	TTGATGCTAA	GGTTTCTGTA	1260
CTGGACTCCC	TGGGTTTAAT	TTGGTGTCT	GTACCCTGAT	TGAGAATGCA	ATGTTTCATG	1320
TAAAGAGAGA	ATCCTGGTCA	TATCTCAAGA	ACTAGATATT	GCTGTAAGAC	AGCCTCTGCT	1380
GCTGCGCTTA	TAGTCTTGTG	TTTGTATGCC	TTTGTCCATT	TCCCTCATGC	TGTGAAAGTT	1440
ATACATGTTT	ATAAAGGTAG	AACGGCATT	TGAAATCAGA	CACTGCACAA	GCAGAGTAGC	1500
CCAACACCAG	GAAGCATTTA	TGAGGAAACG	CCACACAGCA	TGACTTATTT	TCAAGATTGG	1560
CAGGCAGCAA	AATAAATAGT	GTTGGGAGCC	AAGAAAAGAA	TATTTTGCCT	GGTTAAGGGG	1620
CACACTGGAA	TCAGTAGCCC	TTGAGCCATT	AACAGCAGTG	TTCTTCTGGC	AAGTTTTTGA	1680
TTTGTTTATA	AATGTATTCA	CGAGCATTAG	AGATGAACTT	ATAACTAGAC	ATCTGTTGTT	1740
ATCTCTATAG	CTCTGCTTCC	TTCTAAATCA	AACCCATTGT	TGGATGCTCC	CTCTCCATTC	1800
ATAAATAAAT	TTGGCTTGCT	GTATTGGCCA	GGAAAAGAAA	GTATTAAAGT	ATGCATGCAT	1860
GTGCACCAGG	GTGTTATTTA	ACAGAGGTAT	GTAACCTCTAT	AAAAGACTAT	AATTTACAGG	1920
ACACGGAAAT	GTGCACATTT	GTTTACTTTT	TTTCTTCCTT	TTGCTTTGGG	CTTGTGATTT	1980
TGGTTTTTGG	TGTGTTTATG	TCTGTATTTT	GGGGGGTGGG	TAGGTTTAAG	CCATTGCACA	2040
TTCAAGTTGA	ACTAGATTAG	AGTAGACTAG	GCTCATTGGC	CTAGACATTA	TGATTTGAAT	2100



TTGTGTTGTT TAATGCTCCA TCAAGATGTC TAATAAAAGG AATATGGTTG TCAACAGAGA 2160  
CGACAACAAC AACAAA 2176

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 325 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met	Val	Cys	Gly	Ser	Pro	Gly	Gly	Met	Leu	Leu	Leu	Arg	Ala	Gly	Leu	1	5	10	15
Leu	Ala	Leu	Ala	Ala	Leu	Cys	Leu	Leu	Arg	Val	Pro	Gly	Ala	Arg	Ala	20	25	30	
Ala	Ala	Cys	Glu	Pro	Val	Arg	Ile	Pro	Leu	Cys	Lys	Ser	Leu	Pro	Trp	35	40	45	
Asn	Met	Thr	Lys	Met	Pro	Asn	His	Leu	His	His	Ser	Thr	Gln	Ala	Asn	50	55	60	
Ala	Ile	Leu	Ala	Ile	Glu	Gln	Phe	Glu	Gly	Leu	Leu	Gly	Thr	His	Cys	65	70	75	80
Ser	Pro	Asp	Leu	Leu	Phe	Phe	Leu	Cys	Ala	Met	Tyr	Ala	Pro	Ile	Cys	85	90	95	
Thr	Ile	Asp	Phe	Gln	His	Glu	Pro	Ile	Lys	Pro	Cys	Lys	Ser	Val	Cys	100	105	110	
Glu	Arg	Ala	Arg	Gln	Gly	Cys	Glu	Pro	Ile	Leu	Ile	Lys	Tyr	Arg	His	115	120	125	
Ser	Trp	Pro	Glu	Asn	Leu	Ala	Cys	Glu	Glu	Leu	Pro	Val	Tyr	Asp	Arg	130	135	140	
Gly	Val	Cys	Ile	Ser	Pro	Glu	Ala	Ile	Val	Thr	Ala	Asp	Gly	Ala	Asp	145	150	155	160
Phe	Pro	Met	Asp	Ser	Ser	Asn	Gly	Asn	Cys	Arg	Gly	Ala	Ser	Ser	Glu	165	170	175	

Arg Cys Lys Cys Lys Pro Ile Arg Ala Thr Gln Lys Thr Tyr Phe Arg  
 180 185 190  
 Asn Asn Tyr Asn Tyr Val Ile Arg Ala Lys Val Lys Glu Ile Lys Thr  
 195 200 205  
 Lys Cys His Asp Val Thr Ala Val Val Glu Val Lys Glu Ile Leu Lys  
 210 215 220  
 Ser Ser Leu Val Asn Ile Pro Arg Asp Thr Val Asn Leu Tyr Thr Ser  
 225 230 235 240  
 Ser Gly Cys Leu Cys Pro Pro Leu Asn Val Asn Glu Glu Tyr Ile Ile  
 245 250 255  
 Met Gly Tyr Glu Asp Glu Glu Arg Ser Arg Leu Leu Leu Val Glu Gly  
 260 265 270  
 Ser Ile Ala Glu Lys Trp Lys Asp Arg Leu Gly Lys Lys Val Lys Arg  
 275 280 285  
 Trp Asp Met Lys Leu Arg His Leu Gly Leu Ser Lys Ser Asp Ser Ser  
 290 295 300  
 Asn Ser Asp Ser Thr Gln Ser Gln Lys Ser Gly Arg Asn Ser Asn Pro  
 305 310 315 320  
 Arg Gln Ala Arg Asn  
 325

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1893 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGCGGAGCGG GCCTTTTGGC GTCCACTGCG CGGCTGCACC CTGCCCCATC TGCCGGGATC	60
ATGGTCTGCG GCAGCCCGGG AGGGATGCTG CTGCTGCGGG CCGGGCTGCT TGCCCTGGCT	120
GCTCTCTGCC TGCTCCGGGT GCCCGGGGCT CGGGCTGCAG CCTGTGAGCC CGTCCGCATC	180
CCCCTGTGCA AGTCCCTGCC CTGGAACATG ACTAAGATGC CCAACCACCT GCACCACAGC	240

ACTCAGGCCA	ACGCCATCCT	GGCCATCGAG	CAGTTCGAAG	GTCTGCTGGG	CACCCACTGC	300
AGCCCCGATC	TGCTCTTCTT	CCTCTGTGCC	ATGTACGCGC	CCATCTGCAC	CATTGACTTC	360
CAGCACGAGC	CCATCAAGCC	CTGTAAGTCT	GTGTGCGAGC	GGGCCCCGCA	GGGCTGTGAG	420
CCCATACTCA	TCAAGTACCG	CCACTCGTGG	CCGGAGAACC	TGGCCTGCGA	GGAGCTGCCA	480
GTGTACGACA	GGGGCGTGTG	CATCTCTCCC	GAGGCCATCG	TTACTGCGGA	CGGAGCTGAT	540
TTTCCTATGG	ATTCTAGTAA	CGGAAACTGT	AGAGGGGCAA	GCAGTGAACG	CTGTAAATGT	600
AAGCCTATTA	GAGCTACACA	GAAGACCTAT	TTCCGGAACA	ATTACAATA	TGTCATTTCG	660
GCTAAAGTTA	AAGAGATAAA	GACTAAGTGC	CATGATGTGA	CTGCAGTAGT	GGAGGTGAAG	720
GAGATTCTAA	AGTCCTCTCT	GGTAAACATT	CCACGGGACA	CTGTCAACCT	CTATACCAGC	780
TCTGGCTGCC	TCTGCCCTCC	ACTTAATGTT	AATGAGGAAT	ATATCATCAT	GGGCTATGAA	840
GATGAGGAAC	GTTCCAGATT	ACTCTTGGTG	GAAGGCTCTA	TAGCTGAGAA	GTGGAAGGAT	900
CGACTCGGTA	AAAAAGTTAA	GCGCTGGGAT	ATGAAGCTTC	GTCATCTTGG	ACTCAGTAAA	960
AGTGATTCTA	GCAATAGTGA	TTCCACTCAG	AGTCAGAAGT	CTGGCAGGAA	CTCGAACCCC	1020
CGGCAAGCAC	GCAACTAAAT	CCCGAAATAC	AAAAAGTAAC	ACAGTGGACT	TCCTATTAAG	1080
ACTTACTTGC	ATTGCTGGAC	TAGCAAAGGA	AAATTGCACT	ATTGCACATC	ATATTCTATT	1140
GTTTACTATA	AAAATCATGT	GATAACTGAT	TATTACTTCT	GTTTCTCTTT	TGGTTTCTGC	1200
TTCTCTCTTC	TCTCAACCCC	TTTGTAATGG	TTTGGGGGCA	GACTCTTAAG	TATATTGTGA	1260
GTTTTCTATT	TACTAATCA	TGAGAAAAAC	TGTTCTTTTG	CAATAATAAT	AAATTAAACA	1320
TGCTGTTACC	AGAGCCTCTT	TGCTGAGTCT	CCAGATGTTA	ATTTACTTTC	TGCACCCCAA	1380
TTGGGAATGC	AATATTGGAT	GAAAAGAGAG	GTTTCTGGTA	TTCACAGAAA	GCTAGATATG	1440
CCTTAAACA	TACTCTGCCG	ATCTAATTAC	AGCCTTATTT	TTGTATGCCT	TTTGGGCATT	1500
CTCCTCATGC	TTAGAAAGTT	CCAAATGTTT	ATAAAGGTAA	AATGGCAGTT	TGAAGTCAAA	1560
TGTCACATAG	GCAAAGCAAT	CAAGCACCAG	GAAGTGTTTA	TGAGGAAACA	ACACCCAAGA	1620
TGAATTATTT	TTGAGACTGT	CAGGAAGTAA	AATAAATAGG	AGCTTAAGAA	AGAACATTTT	1680
GCCTGATTGA	GAAGCACAAC	TGAAACCAGT	AGCCGCTGGG	GTGTTAATGG	TAGCATTTCT	1740
CTTTTGGCAA	TACATTTGAT	TTGTTCATGA	ATATATTAAT	CAGCATTAGA	GAAATGAATT	1800

ATAACTAGAC ATCTGCTGTT ATCACCATAG TTTTGTTTAA TTTGCTTCCT TTTAAATAAA 1860